Sentiment Analysis on Financial News Articles

Bolanle Onifade

University of Nottingham

Abstract

Over the past few decades, many theories that describe the behaviour of the stock market have been proposed. These theories put forward ideas ranging from those claiming that the stock market cannot be predicted and to those claiming those that with the right assumptions, the stock market can be predicted. Regardless of this, many stock market prediction models (based on technical indicators or non-technical indicators) have been put forward. In this dissertation we aim to consider those non-technical factors that affect stock prices. These non-technical factors come in form of news articles. In the digital age of rapid response to ongoing situations, news articles about events tend to be released as soon as they occur and the assumption is that if we can track news sources, monitor them for the release of relevant news articles, we can use the sentiment orientation of the article (whether positive, negative or neutral) to predict the price of the stock market before the market has a chance to react to the article. This of course poses an important question on how the sentiment orientation of articles. There are several approaches that can be taken towards determining the class of articles but for the purposes of this dissertation, we will focus on the use of Support Vector Machines to perform classification of news articles.

1. Introduction

One can arguably say that state of the world economy has been built with the stock market serving as a base. Hence, while there are other means of evaluating a country's economy, the state of the stock market is perhaps one of the more important means of doing so. Predicting the general direction of stock prices therefore is very important in order to make decisions regarding where and what kind of investment takes place. This work attempts to predict the direction of movement of the (close) stock price of the next working day given information up to and including the current day. We propose a prediction system which incorporates news articles other technical indicators such as the stochastic$\%K$, stochastic $\%D$ (see section 3.6.) .

The advent of the Internet brought with it improvements in many fields ranging from technologies that influence our daily lives to those that may one day take humans to other planets. One of the more mundane improvements is access to news articles – completely changing the way we now make decisions. We no longer have to wait until the following day to find out about events that took place today. Stock market traders are one group of people who rely heavily on this improvement on access to news. Often times, trades are executed using information from the news either consciously and unconsciously. Unconscious decisions can be made because news articles need not necessarily carry extremely significant news in order to be usable to traders – news articles will often bear information about annual earnings, acquisitions and mergers, changes in administration and management as well as stock splits – and thus, news need not bear extremely catastrophic nor positive information, in order to be useful.

Traders will often base their trades on information from news articles regardless of how seemingly important it is. We therefore argue that the key to predicting the stock market is by monitoring extensive sources of news articles and extracting valuable information from the articles which can then be used to predict the stock market. The process of extracting information relating to a person's opinion or emotions automatically from textual data is referred to as sentiment analysis. This work therefore is an interdisciplinary work that ties together sentiment analysis, machine learning and finance for the prediction of the stock market.

2. Literature Review

Bing Liu (2012) highlights three levels of sentiment analysis: document based, sentence based and aspect-entity based. Document-based sentiment analysis pertains to the classification of an entire document and the key assumption is that the entire document focuses on a single entity. Sentence-based analysis is more in-depth analysis of individual sentences. Aspect-entity based analysis focuses on determining the subject of discussion as well as the sentiment. Even further, opinions are split into regular and comparative opinions – regular opinion express opinion on a single entity while comparative opinion expresses opinion on two or more entities.

Bing Liu highlights that the most important indicator of sentiment are sentiment words such as *poor, happy, sad, brilliant* or *excellent.* Of course, one glaring issue with simply analysing based on sentimental words is the use of sarcasm in language. This, in addition with the fact that sentimental words can change orientation (e.g. 'not happy' vs 'very happy') depending on the context means that sentiment analysis can't be reduced simply to a keyword search. It should be noted at this point that financial news articles use sentimental words, making classification harder than with other domains such as movies. Liu discusses extensively sentiment analysis using varying techniques but since we are simply interested in only document-based classification, we focus on articles related to such classification. It is recommended that any reader who wishes to gain a more complete and up-to-date overview of the current sentiment analysis methods should refer to Bing Liu's work on the subject.

One of the more important aspects of document-based classification is the determination of features. Popular ones include terms and their frequency (which is the selected features for this current projected), parts of speech (POS) – introduced by Turney (2002) – and their tags such as sentiment words and sentiment flippers (words that change the orientation of sentiment words such as *not*). Bing Liu also comments that some domains are easier than others, for example movie reviews tend to be easier to analyse than car reviews – this is due to the multi-entity nature of car reviews. Compare:

S1*:* The movie is utterly captivating

S2: While I like the leather seats, the gear is a bit hard to manipulate

In the literature, there are two main approaches to feature generation: a lexicon-based approach and a bag-of-words approach. The lexicon-based approach as used in Whitelaw et al.

(2005) typically involves a set of example words which are then used as a seed for building a larger lexicon through the identification of synonyms in a semi-automatic manner. Bag of words approach typically involves the representation of words as a numerical value indicating their presence, frequency or term frequency-inverse document frequency (TF-IDF) score (Yong, Xu, & Ren, 2011). Please refer to section 3.1.1 for our discussion of TF-IDF. One of the more often cited criticisms of the bag-of-words approach is the loss of semantic association between terms, as the order of words has been lost in the bagging procedure.

Work has been done to improve on the TF-IDF by introducing the delta TF-IDF (Martineau & Finin, 2009). Delta TF-IDF improves the importance placed on words that are unevenly distributed in the corpus and discounts those that are evenly distributed – the rationale being that the less unevenly a feature is represented, the higher the chances that it's relevant for its classification. Delta TF-IDF is formulated mathematically as:

$$\begin{aligned} V_{t,d} &= C_{t,d} * \log_2\left(\frac{|P|}{P_t}\right) - C_{t,d} * \log_2\left(\frac{|N|}{N_t}\right) \\ &= C_{t,d} * \log_2\left(\left(\frac{|P|}{P_t}\right) * \left(\frac{N_t}{|N|}\right)\right) \\ &= C_{t,d} * \log_2\left(\frac{N_t}{P_t}\right) \end{aligned} \quad \textbf{(2.1)}$$

where $C_{t,d}$ is the number of times term t appears in document d, $P_t$is the number of positively labelled documents containing term t, $|P|$ is the number of positively labelled documents, N_t is the number of negatively labelled documents with term t, $|N|$ is the number of negatively labelled documents and $V_{t,d}$ is the value for term t in document d. Using feature values derived by delta TF-IDF on movie review classification, a SVM achieved an accuracy of 88.1% compared to traditional TF-IDF based classification accuracy of 82.85% on the classification of movie reviews.

Feature selection is critical in sentiment analysis as corpuses tend to be polluted with noise, reducing the performance of any developed system. Core techniques involved in feature selection are stop-word removal and stemming. Furthermore, statistical methods such as the point-wise mutual information (Turney & Littman, 2003; Wilson, Wiebe, & Hoffmann, 2005) and chi-square (χ^2) are used in the feature selection process.

The point-wise mutual information $PMI_c(t)$ measures the correlation between a term t and the class c. Assuming mutual independence, PMI can be calculated using the joint distribution and the individual distributions. This is formulated mathematically by Aggarwal and Zhai (2012) as:

$$PMI_c(t) = \log\left(\frac{F(t)\ \cdot p_c(t)}{F(t)\ \cdot P_c}\right) = \log\left(\frac{p_c(t)}{P_c}\right) \quad \textbf{(2.2)}$$

where the expected co-occurrence of c and w is $F(t)\ \cdot P_c$ and the actual co-occurrence is $F(t)\ \cdot p_c(t)$. A PMI of greater than 0 means a positive correlation between w and c and the inverse is the case for a PMI less than 0. We delegate to section 3.1.3 our explanation of χ^2.

Sentiment classification techniques are split into two broad ranges of techniques: machine learning techniques (Yong, Xu, & Ren, 2011; Pang, Lee, & Vaithyanathan, 2002; Kang, Yoo, & Han, 2012; Kaufmann, 2012; Moraes, Valiati, & Neto, 2013), lexicon-based approaches – which is further split into dictionary-based approaches (Guang, Xiaofei, Feng, Yuan, Jiajun, & Chun, 2010; Minging & Bing, 2004; Kim & Hovy, 2004) and corpus-based approaches (Deerwester, Dumais, Landauer, Furnas, & Harshman, 1990). We will focus on the machine learning techniques.

Pang et al. (Pang, Lee, & Vaithyanathan, 2002) used a corpus of reviews that were classified according to the number of stars associated with the text reviews and predicted these ratings using naïve bayes, maximum entropy and Support Vector Machines. They compared the performance of unigrams, unigrams + bigrams, bigrams, unigrams + POS, adjectives, top unigrams and unigrams + positions using three-fold cross-validation. The performance of each variation in features depended heavily on the type of machine learning technique used. SVMs on average performed better than both naïve bayes and maximum entropy. It's important to note that the classifiers performed best when the terms were represented simply as a presence as opposed to frequency. Unigrams performed best in their experiments using the Support Vector Machine-based model with an accuracy of 82.9%.

Yong et al. (2011) propose a method of sentiment analysis similar to the approach this work has taken, pre-processing the words by removing stop words and tokenisation. Words are then sorted based on calculated mutual information and a number of words are selected as the feature items and classified using an SVM. They neglect to specify the domain and source of the corpus, making comparison with this work difficult. However, they achieve very high classification rates with total precision of 81.11%, recall of 81.42% and F-value of 81.25 in a closed test.

3. Background Knowledge

3.1. Sentiment Analysis

In order to make the documents useful to numeric classifiers such as the neural network or support vector machine, the terms need to be transformed into their weighted forms. The weighting of the terms is shown to be at least as important as their selection (Strzalkowski, 1994).

The vector space model introduced by Salton et al. (1975) is the preferred algebraic method of representing textual documents. The document is represented as a single vector where each dimension in the vector is a single feature. Features that do not exist in a particular document d are simply given a value of 0. The weights are usually calculated using simply their presence (binary), counts (integer), frequency(float) or their term frequency inverse document frequency (TF-IDF) score (float). Hence we can say that in a corpus of $|D|$ documents and m features, a document d_j is represented as:

$$d_j = \left(w_{1,j}, w_{2,j}, w_{3,j}, \dots, w_{m,j}\right) \tag{3.1}$$

where $1 \leq j \leq |D|$. We described an improvement on TF-IDF in chapter 2 and now, we provide the mathematics behind the traditional TF-IDF (Salton, Wong, & Yang, 1975). Given, the document$j$, the weight $w_{t,d}$ is calculated as:

$$w_{t,d} = tf_{t,d} \cdot \log\left(\frac{|D|}{1 + |\{d' \epsilon D|\ t \epsilon d'\}|}\right) \quad \textbf{(3.2)}$$

where $|\{d' \epsilon D|\ t \epsilon d'\}|$ is the number focuments which contain the term t. 1 is added to prevent divide-by-zero errors. $tf_{t,d}$ is the term frequency and is given by the formula:

$$tf_{t,d} = \frac{n_{t,d}}{\sum_m n_{m,d}} \quad \textbf{(3.3)}$$

where $n_{t,d}$ is the number of times feature f_t occurs in document d_j and $\sum_m n_{m,d}$ is the total number of terms in the document d_j. The higher the value assigned to a feature, the importance, it is given.

3.2. Support Vector Machines

Techniques in machine learning generally are classified into two main divisions: supervised and non-supervised learning. Support Vector Machines (SVMs) are classified under the former. This means that the process by which SVMs solve tasks is by first undergoing a training phase in which input data (referred to as a set of examples or instances) are used to learn a model that can then be used to classify new instances into a category or a set of categories. SVMs typically solve binary classification[1] problems.

SVMs work by projecting and mapping (by using a kernel function) training data instances into a high-dimensional feature space, such that the gap between the two categories is maximised. Hence, new instances can be classified by side of the gap the point falls on. *Gap* is a vague term and is difficult to define. Instead, SVMs work by constructing a hyperplane that separates the two categories – this is referred to as the maximum margin hyperplane. The maximum margin hyperplane is sought after theoretically, as such an hyperplane should lead to the lowest generalisation error.

A hyperplane which defines the decision boundary of the classes is in turn defined in terms of a set of points whose set of points x satisfy the equation.

$$w^T \cdot \phi(x) + b = 0$$

where w^T is the weight vector of the hyperplane, $\phi(x)$ is the kernel method that maps the data points to a (hopefully)linearly feature space and b is the bias.

For linearly separable data, two hyperplanes can be defined which completely separate the data such that there are no points between them. The area in space that is bordered by the two hyperplanes is called a margin. The two hyperplanes are defined by the following equations:

$$H_1: \quad w^T \cdot \phi(x) + b_0 = +1$$
$$H_2: \quad w^T \cdot \phi(x) + b_0 = -1$$

[1] Binary classification involves the training of a model to differentiate between only two classes.

In order to define H_1 and H_2, vectors must be selected which just touch the margins – these vectors are referred to as support vectors as these are the only data instances required to represent the model. The margin can also be defined more formally in terms of d_1 and d_2 which are the shortest distance between the maximum margin hyperplane H_0 and the closest positive support vector and the shortest distance between the H_0 and the closest negative support vector, respectively.

In figure 3.1., $\frac{b}{||w||}$ is defined as the offset between the maximum margin hyperplane and the origin. We can thus say that $d = d_1 + d_2$ is the width between H_1 and H_2. Given that the distance between H_1 and H_0 is $\frac{|w^T \cdot \phi(x)|}{||w||} = \frac{1}{||w||}$, the total distance d is then $\frac{2}{||w||}$. We therefore need to minimise $||w||$ in order to maximise the margin, d. This problem can thus be formulated mathematically as:

$$y_i(w^T \cdot \phi(x_i) + b) \geq 1 = \begin{cases} w^T \cdot \phi(x_i) + b \geq +1 \; when \; y_i = +1 \\ w^T \cdot \phi(x_i) + b \leq -1 \; when \; y_i = -1 \end{cases} \quad \textbf{(3.4)}$$

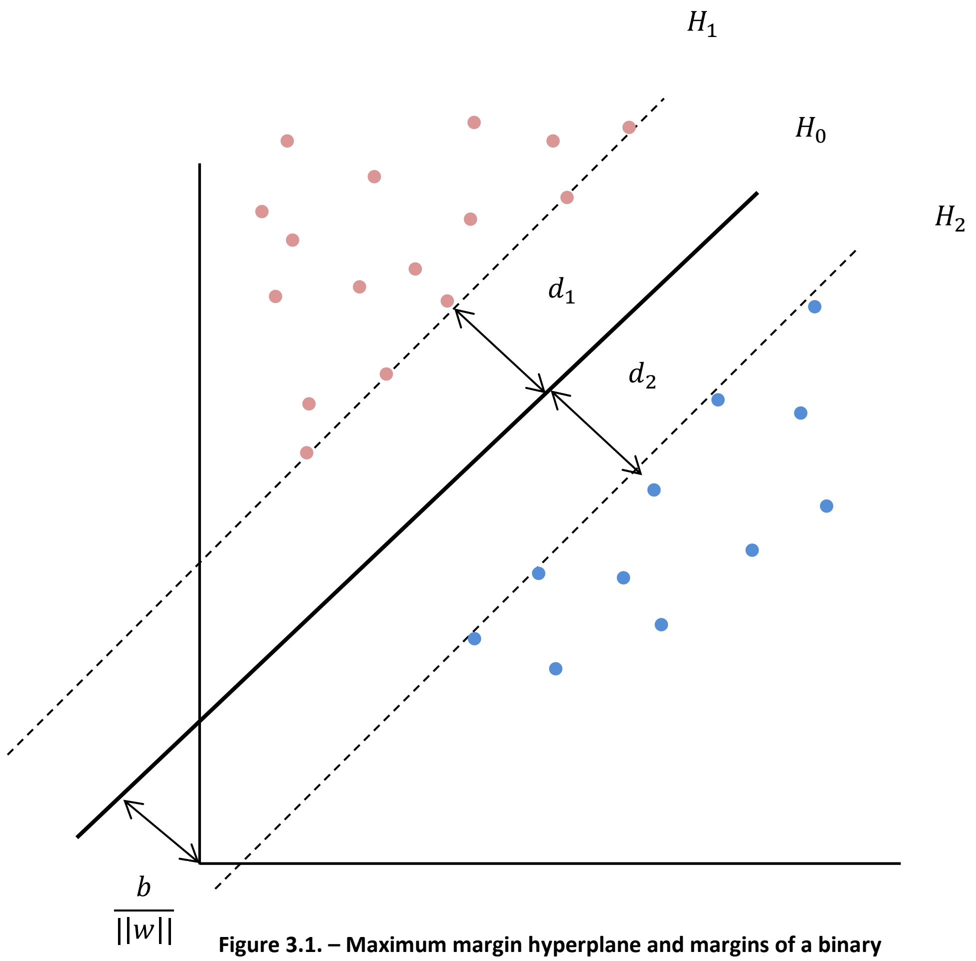


Figure 3.1. – Maximum margin hyperplane and margins of a binary classification SVM-based model

where x_i is an example, y_i is the target class. Minimising $||w||$ can be reduced to minimising $\frac{1}{2}||w||^2$

Having formulated the linearly separable case, we note that not all SVM-applicable problems are linearly-separable; in fact, most of them are not, even in kernel or feature space. For this, we introduce a slack variable $\xi_n \geq 0$ which acts as a penalty term for misclassified examples. ξ_n is thus formulated mathematically as:

$$\xi_n = \begin{cases} 0, & \text{if correctly classified} \\ |t_i - y_i(\phi(x_i) \cdot w^T + b)|, & \text{otherwise} \end{cases} \tag{3.5}$$

Hence, $y_i(w^T \cdot \phi(x_i) + b) \geq 1$ can be rewritten as $y_i(w^T \cdot \phi(x_i) + b) \geq 1 - \xi_n$. This new formulation is referred to as a soft margin and the modified optimisation goal is:

$$arg \max_{\xi_n, w} c \sum_{n=1}^{N} \xi_n + \frac{1}{2} ||w||^2 \tag{3.6}$$

Often, we find that there is more than a single category to classify. There are two main approaches to multiclass classification problems: one-versus-one and one-versus-rest. One-versus-one attempts to differentiate between each two pairs of class while one-versus-many differentiates between a single class and the other classes.

We conclude our discussion of Support Vector Machines with the kernel function, ϕ. The resulting feature space is heavily dependent on the exact function mapping and there are a few popular kernel functions, which are simply listed here.

a. Linear kernel is defined as $k(x, x') = x^T x'$ (i.e. no mapping)
b. Polynomial kernel is defined as $k(x, x') = (x^T x' + c)^d$ where c is a constant which accounts for the influence of higher-order terms versus the lower-order terms d is the degree
c. Radial basis function kernel $k(x, x') = \exp\left(-\gamma ||x - x'||^2\right)$ where γ is a hyper-parameter referred to as the kernel bandwidth.
d. Gaussian kernel: $k(x, x') = \exp(-\frac{||x-x'||^2}{2\sigma^2})$

4. Scientific Method

The proposed system is comprised of two main parts: sentiment-based news classification and price prediction. Any developed system cannot completely rely on news as news is sporadic and there's no guarantee that news regarding a certain entity will be released for every single trading day and thus, there needs to be a means by which prediction can still take place. This comes in the form of technical data. Hence, there are two phases of classification – the classification of pure news articles and the classification of technical data. Figure 4.1 shows the system overview, ignoring news labelling.

Unlike sentiment classification for domains such as films, cars or music, happiness, sadness and ambivalent news articles may not bear much information about the progress of the entity. While the sentiment of the article is what we aim to extract, a cursory look at any news article that bears financial information will show that news articles aren't very sentimental. This indicates that classification based simply on human sentiment while might be accurate might not be as successful given that we aim to predict the stock market. Therefore, to supplement classification based on sentiment (which we refer to emotion sentiment with categories *happy, sad, neutral*), we also classify based on the progression of the company. This means that the articles get classified into an additional set of categories (positive, negative and neutral).The aim with the progress classification is that articles get classified based on what the news article's evaluator expects as regards to whether the entity's stock price will go up, down or simply stay the same as a result of the article.

4.1. Data Acquisition

4.1.1. News Articles Acquisition

It's clear that the very first task to be performed is the acquisition of news articles whether labelled or non-labelled. Although a fair bit of work have been done using this particular approach to stock price prediction, we were unable to find any publicly available datasets that fell in line with the purposes of this dissertation. Hence, a dataset was generated from online news sources.

Although selecting news sources seems like a trivial task, it requires careful consideration as the news sources has to be able to satisfy the following requirements:

i. Has to be popularly read, especially by traders. This is particularly important because a high level of trust needs to be placed in the news source, enough to determine that significant changes in stock price trend will be reflected in the news articles.
ii. The news sources should have decent coverage of news– to ensure that we gather as much data as possible.

Given these requirements, investors on online forums (as well as individuals with knowledge of finance) were asked which news sources were read and the following sources were given: Reuters, Bloomberg, Financial Times, Market Watch, Yahoo Finance.

The table below shows the number of articles that were gathered for each company.

Company Name	Number of Articles
Chevron	88
Cocacola	52
Disney	108
Exxon	120
Goldman	731
IBM	118
JP Morgan	613
Microsoft	259
Pfizer	115
Visa	48

Figure 4.1 – Number of articles collected for each company

4.1.2. Stock Data Acquisition

The price daily values were collected for the period from the 1st of January, 2013 to the 30th of September, 2014. Of course, the stock price data is only released for working days so this accounts for only 440 working days.

4.2. Manual Labelling

Manual labelling of data is simply reading each news article and labelling them by hand. The evaluators are asked to estimate the company's progression based on the news article. Their estimates can fall into the following categories: (up, down, neutral). The evaluators are also asked to provide the sentiment of the article (happy, sad, neutral). From henceforth, for clarity purposes, we shall refer to the former as "progress sentiment" and the latter as "emotion sentiment"

One might think that there is a perfect correlation between the two sets of categories. However, there can be differences between the two. For illustrative purposes, we will examine a couple of cases in which there are differences: The headline "Exxon Mobil reports Fire, oil spill at Nigerian terminal", evokes a feeling of sadness but due to the established nature of Exxon, it's unlikely that this event is going to lead to a massive dent in the stock price, we give the article a progress sentiment of neutral (because the article doesn't go on to indicate that Exxon will suffer from this incident). Another article that wouldn't be expected to change the stock price much is "JP Morgan employee falls to death from building roof in Hong Kong". It's however clear that the article is "sad", but the effect on JP Morgan's progress would virtually be nil.

If the previous two examples give the impression that from headlines, we can always tell the emotion sentiment of an article, it would be wrong. In fact, a seemingly neutral headline such as "Coca-cola names Walter Finance Chief as Fayard Retires" goes on to discuss the recent struggles of coca-cola, therefore giving it a emotion sentiment of sad and a progress sentiment of neutral. In the same strain, we discovered articles can both be up for progress sentiment and emotion sentiment; this would be the case for articles that discuss an entity's growing business.

We finish off this section by pointing out that the more similar the correlation values are between the projected trend line (generated via piecewise linear approximation) and the actual stock price trend line, the higher the similarity is between the two trends.

4.3. Data Pre-processing

Retrieved news articles are in HTML format. The news articles therefore need to be converted into plain text, tokenised (into unigrams and bigrams), stemmed and have stop words removed (discussed in chapter 3) before classification can take place.

4.4. Document Representation

After the completion of all pre-processing steps, the documents are now ready to be transformed into vectors. The library Scikit-learn provides the TfidfVectorizer that converts the news articles into a TF-IDF-weighted document-term matrix. TF-IDF has been discussed in (chapter 3).

4.5. Feature Selection or Reduction

In the literature, the chi-squared method for feature selection and the SVD method are popular and but we decided to use χ^2 as SVD is a very expensive method.

5. Classification

The final step is to train the SVM to predict the news article. In order to truly evaluate the SVM, cross validation over the data set was performed. 10-fold cross validation ensured we got the performance of the hyper-parameters of the SVM. The results of each fold were evaluated using the following metrics: confusion matrix, recall, precision and f-measure. These results can then be averaged over the number of folds to determine the overall performance of the hyper-parameters.

6. Experiments and Results

6.1.1. Progress Sentiment Classification

In order to set the weights, we need to look at the support for each class. Figure 5.18 shows the number of articles supporting each class. The hyper-parameters used for configuring the SVM are as follows:

LinearSVC (the class used for classification) implements the one-versus-rest classifier for multiclass problems. We use a *C* value of 2.9. We use automatically set class weights (which modifies the C-values for each class) for the SVM as Figure 5.18 shows, the classes are not represented equally in the training sets. Using a StratifiedKFold cross validator, we can preserve the percentage of representation for each sample.

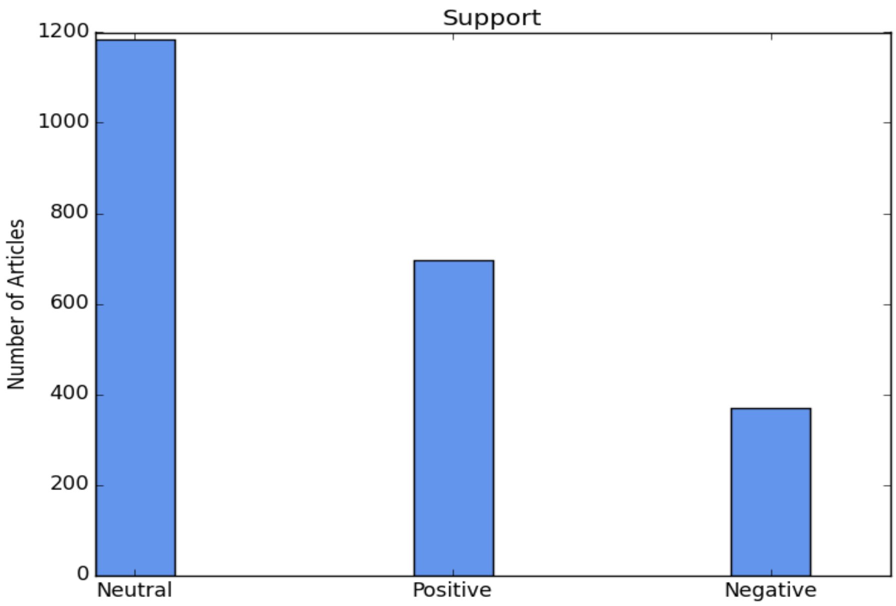


Figure 6.1 – support for the various classes (manual/progress)

Similar settings were used for bigrams, unigrams and combination experiments. In order to determine accuracy, we use cross validation and the following metrics: f-measure, recall, precision and confusion matrix. We compute the average of the scores of all the folds. The confusion matrixes for unigram, bigram and combination (Figure 5.19) show an overview of the accuracy for the three classes.

The confusion matrixes show that there aren't very big differences in the performances of the three methods of tokenisation (except when classifying positive articles). It's very difficult to explain why this is the case except that all three methods carry similar levels of information for this problem.

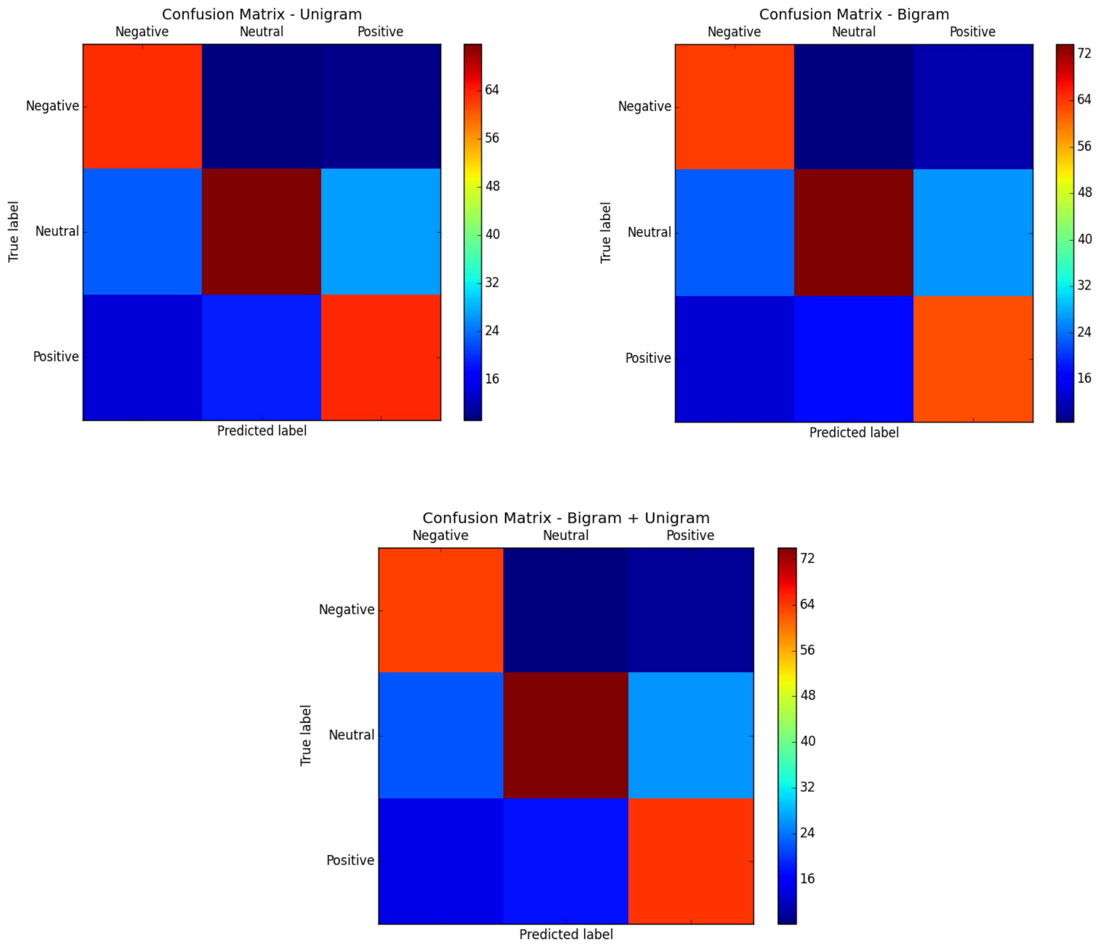


Figure 6.2 – Confusion matrices (Manual/Progress). Top left – Unigram, Top Right – Bigram, Bottom – Unigram + Bigram in percentages

	Unigram	Bigram	Unigram + Bigram
F-measure	68.62	69.82	70.51
Recall	68.68	69.97	70.58
Precision	69.04	70.17	70.62

Figure 6.3 – Table of performance of linear SVM measured by cross validation (manual/progress)

Delving into the actual numbers, we see that overall, the bigram does better than the unigram and the combination of both does better than either of them singularly. Combining this information with the confusion matrix, we see that bigrams and the combination perform better due to being able to slightly classify positive news articles better.

Given the similarities in the values, T-tests were performed (with an alpha of 0.05 and a $n - 1$ degree of freedom) to determine there are any significant differences between the results attained with the features. The t-tests confirmed that there are no statistically significant differences between the results.

6.1.2. Emotion sentiment Classification

Poorer results were achieved for the classification of emotion sentiment in general. This is contradictory to the initial belief that emotion sentiment would be easier than progress sentiment to classify. We performed classification using a linear SVM as before. The settings for emotion sentiment were quite different. In addition, classification performance for the emotion sentiment was quite poor overall. As per the previous section, we start by introducing the frequencies for the classes (Figure 5.21)

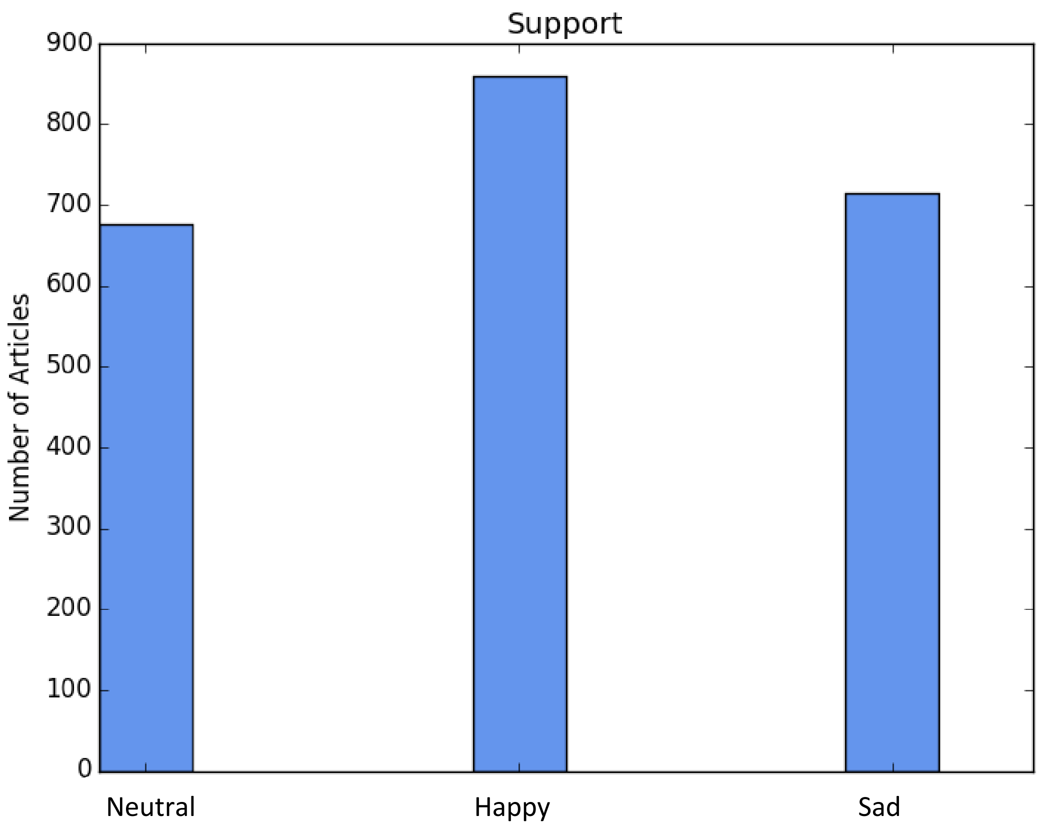
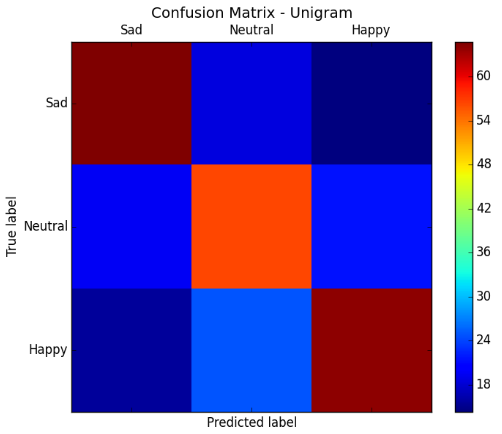
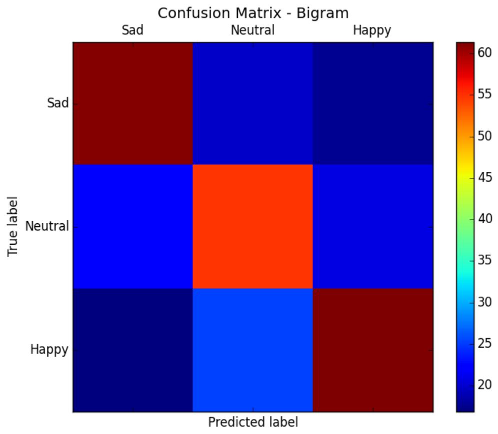


Figure 6.4 – Support for the classes (Manual/ Emotion)

Different hyper-parameter settings were used for classification. The C parameter was set to higher levels with a value of $1 * 10^3$. The other parameters, such as the class weight were also set automatically based on the class.

Considering the confusion matrices (Figure 5.15), we see that the all three methods of tokenising perform very similarly as before. A possible reason for this is that news articles often bear mixed feelings. On the surface, it may seem that news articles bear emotion sentiment orientations that lean towards one way or the other but this isn't so. News articles often carry information that lean to both sides. A classic example of such news articles is articles that discuss "happy" sentiment. In a few of these articles, there's also discussion of past "sad" sentiment that led to perhaps structural changes that result in improvement. Hence, while progress sentiment might be relatively clear, emotion sentiment can often be ambiguous when it comes to classifying neutral articles.





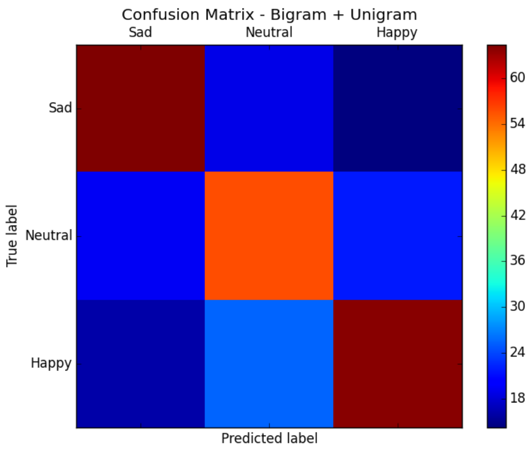


Figure 6.5 – Confusion matrices (Manual/ Emotion). Top left – Unigram, Top Right – Bigram, Bottom – Unigram + Bigram in Percentages

	Unigram	Bigram	Unigram + Bigram
F-measure	62.00	60.87	63.68
Recall	62.48	61.06	63.92
Precision	62.23	61.15	63.99

Figure 6.6 - Table of performance of linear SVM measured by cross validation (Manual/ Emotion)

Bigrams are typically expected to do better than unigram due to the fact they retain sentence structure but clearly, bigrams doesn't do very well for this problem looking at the performance measures in Figure 5.16. However, combination of both performs better than either but not by much.

7. Conclusion

The results of the experiments show that further work on the dataset might lead to promising and useful results – especially in regards to actual stock price prediction. This work further contributes to the growing literature on sentiment analysis but evaluates on a relativity unexplored domain and introduced two measures by which financial news articles might be evaluated – both in terms of emotion and progression.